

# Empirical central limit theorem for cluster functionals without mixing.\*

PAUL DOUKHAN<sup>†</sup> AND JOSÉ-GREGORIO GÓMEZ<sup>‡</sup>

## Abstract

We prove central limit theorems (CLT) for empirical processes of extreme values cluster functionals as in Drees and Rootzén (2010). We use coupling properties enlightened for Dedecker & Prieur's  $\tau$ -dependence coefficients in order to improve the conditions of dependence and continue to obtain these CLT. The assumptions are precisely set for particular processes and cluster functionals of interest. The number of excesses provides a complete example of a cluster functional for a simple non-mixing model (AR(1)-process) for which ours results are definitely needed. We also give the expression explicit the covariance structure of limit Gaussian process.

Also we include in this paper some results of Drees (2011) for the extremal index and some simulations for this index to demonstrate the accuracy of this technique.

*Keywords and phrases:* Extremes, clustering of extremes, cluster functional of extremes, extremal index, uniform central limit theorem,  $\tau$ -weak dependence, tail empirical process.

## 1. INTRODUCTION

Drees & Rootzén [2010]'s scheme prove limit theorems for empirical process of cluster functionals (EPCF). In statistics, Gómez [2015] proves a CLT in finite dimensional convergence (fidis) under weaker conditions, and considers an example for which a functional CLT is not necessary.

We extend the result under  $\tau$ -weak dependence. The classical example of a non-mixing autoregressive model demonstrates the importance of this functional extension. Moreover the estimation of the extremal index provides us with a suitable example of application of the functional CLT.

For a real-valued random process  $(X_i)_{i \in \mathbb{N}}$ , a typical example of a EPCF is the tail empirical process:

$$T_n(x) = \frac{1}{\sqrt{nv_n}} \sum_{i=1}^n (\mathbb{1}\{X_i > a_n x + u_n\} - \mathbb{P}(X_1 > a_n x + u_n)), \quad x \geq 0, \quad (1)$$

\*This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

<sup>†</sup>Institut Universitaire de France (IUF) and <sup>‡</sup>.

<sup>‡</sup>Département de Mathématiques, Université de Cergy - Pontoise. 95000 Cergy - Pontoise, France.

<sup>0</sup>E-mail addresses: doukhan@u-cergy.fr and jose.gomez@u-cergy.fr

<sup>0</sup>AMS 2000 Subject Classifications: Primary 60G70; secondary 60F05, 60F17, 62G32.

where  $v_n = \mathbb{P}\{X_{n,1} \neq 0\}$  is decreasing to zero,  $(u_n)_{n \in \mathbb{N}}$  is a non-decreasing sequence of thresholds and  $(a_n)_{n \in \mathbb{N}}$  is a sequence of positive constants. This process has been considered by Drees and Rootzén under suitable dependence conditions (in particular under  $\alpha$  and  $\beta$ -mixing conditions), where they prove its uniform convergence to a Gaussian process  $T$  under additional conditions. For example, they prove the convergence of this tail empirical process for the cases of  $k$ -dependent sequences or stable AR(1)-processes [Rootzén, 1995], ARCH(1)-processes [Drees, 2002, 2003] and some applications for solutions of stochastic difference equations [Drees, 2000, 2002, 2003]. Finally, they use the cluster functionals setting in [Yun, 2000] and [Segers, 2003] to generalize such empirical processes under  $\beta$ -mixing in [Drees & Rootzén, 2010].

Unfortunately, the mixing processes family is very restrictive. This can be noted with the following AR(1)-process, solution of the recursion:

$$X_k = \frac{1}{b}(X_{k-1} + \xi_k), \quad k \in \mathbb{Z}, \quad (2)$$

where  $b \geq 2$  is an integer and  $(\xi_k)_{k \in \mathbb{N}}$  are independent and uniformly distributed random variables on the set  $U(b) := \{0, 1, \dots, b-1\}$  which is not even  $\alpha$ -mixing, as this is shown in [Andrews, 1984] for  $b = 2$  and in [Ango Nze & Doukhan, 2004] for  $b > 2$ . Thus, *the results in [Drees & Rootzén, 2010] can not be used here!* However, such process (2) is  $\tau$ -weakly dependent as is shown in [Dedecker & Prieur, 2004a]. The same situation happens in a general way for the causal Bernoulli shifts, Markovian models, etc. This is thus useful to improve on the CLT for empirical processes of extreme cluster functionals proposed by Drees & Rootzén [2010] for more general classes of weakly dependent processes.

In order to do this, we use of the coupling results of Dedecker & Prieur [2004a, 2005] under  $\tau$ -dependence assumptions to use Van Der Vaart & Wellner [1996]'s results of tightness and asymptotic equicontinuity (under independence) together with the fidis convergence of the EPCF.

This paper is organized as follows. In Section 2, we recall basic definitions and notations: cluster functionals, the triangular arrays (or normalized random variables excesses) and examples. Then, we give the definition of cluster functionals empirical processes and close the section with a simple version of the CLT of those empirical processes. In Section 3 we begin with the fidis convergence of the EPCF, followed by the conditions to obtain asymptotic tightness and asymptotic equicontinuity of those processes to obtain uniform convergence. We close this section with an application: block estimator of the extremal index. In Section 4 we develop an example similar to (1) for the multidimensional case for the case of the AR(1)-inputs (2). Also a simulation study for the extremal index to demonstrate the accuracy of this technique. The  $\tau$ -weak dependence with some examples and the proofs are reported in Appendix.

## 2. BASIC DEFINITIONS AND NOTATIONS

To define the empirical processes of cluster functionals it is necessary to consider two important ingredients: the cluster functionals and the excesses over high thresholds.

## 2.1. Cluster Functionals

Let  $(E, \mathcal{E})$  be a measurable subspace of  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  for some  $d \geq 1$  such that  $0 \in E$ . Following the deterministic definition of Drees & Rootzén [2010]<sup>1</sup>, we consider the set of  $E$ -valued sequences of finite length, *i.e.*,

$$E_{\cup} := \{(x_1, \dots, x_r) : x_i \in E \quad \forall i = 1, \dots, r; \quad \forall r \in \mathbb{N}\},$$

equipped with the  $\sigma$ -field  $\mathcal{E}_{\cup}$  induced by Borel- $\sigma$ -fields on  $E^r$ , for  $r \in \mathbb{N}$ . Let  $x \in E_{\cup}$ , then we can write  $x = (x_1, \dots, x_r)$  for some  $r \in \mathbb{N}$ . The **core**<sup>2</sup>  $x^c \in E_{\cup}$  of  $x$  is defined by

$$x^c := \begin{cases} (x_{r_I}, x_{r_I+1}, \dots, x_{r_S}), & \text{if } x \neq 0_r \text{ (the null element in } E^r) \\ 0, & \text{otherwise} \end{cases}$$

where  $r_I := \min\{i \in \{1, \dots, r\} : x_i \neq 0\}$  (first non-null value of the block  $x$ ) and  $r_S := \max\{i \in \{1, \dots, r\} : x_i \neq 0\}$  (last non-null value of the block  $x$ ). A **cluster functional** is a measurable map  $f : (E_{\cup}, \mathcal{E}_{\cup}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  such that

$$f(x) = f(x^c), \quad \text{for all } x \in E_{\cup}, \quad \text{and } f(0_r) = 0 \quad (\forall r \geq 1). \quad (3)$$

Under the properties (3), it is easy to build a large amount of examples of cluster functionals. Nevertheless, the typical examples used to build estimators through these cluster functionals are functionals of the type:

$$f(x_1, \dots, x_r) = \sum_{i=1}^r \phi(x_i), \quad (4)$$

where  $\phi : E \rightarrow \mathbb{R}$  is a measurable function such that  $\phi(0) = 0$ . Generally speaking, these functions  $\phi$  are indicator functions (or functions which are product of another measurable function  $H : E \rightarrow \mathbb{R}$  with an indicator function). Another classic example is the component-wise maximum of a cluster:

$$f(x_1, \dots, x_r) = \max_{1 \leq i \leq r} x_i, \quad (5)$$

for  $E = [0, \infty)$ .

Under the set  $E = [0, \infty)$ , two particular examples that we shall only mention here with a view to motivating further work, are the following functionals:

- Balanced periods at  $u > 0$ ,

$$f(x_1, \dots, x_r) = \begin{cases} \mathbb{1}\{\sum_{i=1}^r (x_i - u) \mathbb{1}\{x_i > 0\} = 0\}, & \text{if } x_i > 0 \text{ for some } i = 1, \dots, r \\ 0, & \text{if } x_i = 0 \text{ for all } i = 1, \dots, r. \end{cases}$$

<sup>1</sup>This definition is given by Yun [2000] and Segers [2003], for the real case

<sup>2</sup>Note that the core also considers the null values that exist between the non-null values. For example.  $(0, 1, 2, 4, 0, 3, 0, 1, 0, 0)^c = (1, 2, 4, 0, 3, 0, 1)$ , which is the smaller sub-block of  $x = (0, 1, 2, 4, 0, 3, 0, 1, 0, 0)$  which contains all non-null values as well as the null values between them.

- Maximum sum (greater than the level  $u > 0$ ) of consecutive excesses,

$$f(x_1, \dots, x_r) = \max_{1 \leq p < q \leq r} \left( \sum_{y \in H_{p,q}} y \right) \mathbb{1} \left\{ \sum_{y \in H_{p,q}} y > u \right\},$$

with the notation:  $H_{p,q} = \{x_p, x_{p+1}, \dots, x_q\} \subseteq \{x_1, \dots, x_r\}$  such that  $y > 0$ ,  $\forall y \in H_{p,q}$ , where  $p, q \in \{1, 2, \dots, r\}$ .

## 2.2. The excesses and their normalizations

Without going yet into formalities, first let us consider the following examples that motivate the use of the triangular arrays  $(X_{n,i})_{1 \leq i \leq n, n \in \mathbb{N}}$  throughout this work.

1.- Let  $(X_i)_{i \in \mathbb{N}}$  be a real-valued stationary random process with marginal cumulative distribution function  $F$  and let  $(u_n)_{n \in \mathbb{N}}$  be a non-decreasing sequence of thresholds such that  $u_n \uparrow x_F$ , where

$$x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}, \quad v_n = \mathbb{P}\{X_1 > u_n\} \xrightarrow{n \rightarrow \infty} 0.$$

If we want to study the process  $X'_{n,i} = X_i - u_n | X_1 > u_n$ , first observe that the tail distribution function of  $X_i$  may be asymptotically degenerated as  $n \rightarrow \infty$ , which means that there exists a point  $a \in \mathbb{R}$  such that

$$\bar{P}_n(x) = \mathbb{P}\{X_1 - u_n > x | X_1 > u_n\} \xrightarrow{n \rightarrow \infty} \mathbb{1}\{x \leq a\}.$$

However, if  $F$  belongs to the domain of attraction of some extreme-value distribution, then by a result in [Pickands, 1975], there exists  $\gamma \in \mathbb{R}$  and a sequence of positive constants  $(a_n)_{n \in \mathbb{N}}$  (depending on the sequence  $u_n$ ) such that

$$P_n(x) = \mathbb{P}\{X_{n,1} > x | X_1 > u_n\} \xrightarrow{n \rightarrow \infty} \begin{cases} (1 + \gamma x)_+^{-1/\gamma}, & \text{if } \gamma \neq 0 \\ e^{-x}, & \text{if } \gamma = 0 \end{cases}$$

locally uniform in  $(0, \infty)$ , where

$$X_{n,i} = \left( \frac{X_i - u_n}{a_n} \right)_+ := \max \left\{ \frac{X_i - u_n}{a_n}, 0 \right\}, \quad \text{for } 1 \leq i \leq n; \quad (6)$$

are the normalized excesses of  $X_i$  over  $u_n$ .

2.- For the multidimensional case we may consider the following example. For  $d \geq 1$ , let  $(\mathbf{X}_i)_{i \in \mathbb{N}}$  be a  $\mathbb{R}^d$ -valued random process such that  $\mathbf{X}_i = (X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(d)})$  admits coordinates with the same marginal distribution, then in this case, a standardization of  $\mathbf{X}_i$  is:

$$X_{n,i} = \left( \left( \frac{X_i^{(1)} - u_n}{a_n} \right)_+, \left( \frac{X_i^{(2)} - u_n}{a_n} \right)_+, \dots, \left( \frac{X_i^{(d)} - u_n}{a_n} \right)_+ \right), \quad (7)$$

where  $(u_n)_{n \in \mathbb{N}}$  and  $(a_n)_{n \in \mathbb{N}}$  are defined as in eqn. (6). Here,  $X_{n,i}$  is the vector of normalized excesses over the threshold  $u_n$  for each coordinate.

Another interesting example is the normalization of  $d$  consecutive excesses of real-valued random variables  $(X_i)_{i \in \mathbb{N}}$ , i.e.

$$X_{n,i} = \left( \left( \frac{X_i - u_n}{a_n} \right)_+, \left( \frac{X_{i+1} - u_n}{a_n} \right)_+, \dots, \left( \frac{X_{i+d-1} - u_n}{a_n} \right)_+ \right). \quad (8)$$

This example is given in Section 3 - [Drees & Rootzén, 2010]. Observe that this example is a particular case of the example of eqn. (7) which corresponds to  $X_i^{(j)} = X_{i+j-1}$  for  $1 \leq j \leq d$ .)

Notice that this example brings information on the extremal dependence structure. Some applications of this standardization could be:

- (i)  $d$  consecutive days of rain are observed in a given city, such that the volume of precipitations may be larger than the volume of water that can be drained (through sewers, soil, rivers, etc.),
- (ii)  $d$  very large claims are reported to an insurance company in a very small time interval (with respect to typical cases) which this can be a risk with respect to the response capacity of the insurance company, and
- (iii)  $d$  consecutive days of low temperatures observed in a given city, such that the power consumption (due to heating, etc.) endangers the response capacity of the company in charge of the energy distribution.

In a general way, let  $(E, \mathcal{E})$  be a measurable subspace of  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  for some  $d \geq 1$  such that  $0 \in E$ . We denote by  $(X_{n,i})_{1 \leq i \leq n}$  as the  $E$ -valued row-wise stationary standardized random variables, defined on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ , which are built from a stationary random process  $(X_i)_{i \in \mathbb{Z}}$ , in a way such that the standardization  $X_{n,i}$  maps all "non-extreme" values to zero. Additionally, it should satisfy that the sequence of conditional distributions of  $X_{n,1}$  given that  $X_{n,1}$  belongs of the failure set  $S \subseteq E \setminus \{0\}$  (i.e.  $P_n(\cdot | S) := \mathbb{P}\{X_{n,1} \in \cdot | X_{n,1} \in S\}$ ), converge weakly to some non-degenerate limit.

### 2.3. Empirical Processes of Cluster Functionals

Now, we want to apply cluster functionals to blocks of  $E$ -valued random variables excesses over a determined thresholds sequence and to define the empirical process indexed by these functionals.

In order to do that, first let us consider a row-wise stationary  $E$ -valued triangular array  $(X_{n,i})_{1 \leq i \leq n, n \in \mathbb{N}}$ , defined on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

Let  $Y_{n,j}$  be the  $j$ -th block of  $r_n$  consecutive values of the  $n$ -th row of  $(X_{n,i})$ . That is, we have  $m_n := \lfloor n/r_n \rfloor = \max\{j \in \mathbb{N} : j \leq n/r_n\}$  blocks

$$Y_{n,j} := (X_{n,i})_{(j-1)r_n+1 \leq i \leq jr_n} \quad (9)$$

of length  $r_n$ , with  $1 \leq j \leq m_n$ . In order to simplify future notations, since  $(X_{n,i})_{1 \leq i \leq n}$  is stationary for each  $n$ , then we can denote by  $Y_n$  to the "generic block" such that  $Y_n \stackrel{D}{=} Y_{n,1}$ . Let  $\mathcal{F}$  be a class of cluster functionals. The **"empirical process  $Z_n$  of cluster functionals"**

in  $\mathcal{F}$ , is the process  $(Z_n(f))_{f \in \mathcal{F}}$  defined by

$$Z_n(f) := \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{m_n} (f(Y_{n,j}) - \mathbb{E}f(Y_{n,j})), \quad (10)$$

where  $v_n := \mathbb{P}\{X_{n,1} \in S\}$  and  $S$  is the failure set.

In order to begin approaching the convergence in fidis of the EPCF (10), observe that if the blocks  $(Y_{n,j})_{1 \leq j \leq m_n, n \in \mathbb{N}}$  are independents and if we take in account the following essential convergence assumptions:

$$\text{(C.1)} \quad \mathbb{E} \left[ (f(Y_n) - \mathbb{E}f(Y_n))^2 \mathbb{1}_{\{|f(Y_n) - \mathbb{E}f(Y_n)| > \epsilon \sqrt{nv_n}\}} \right] = o(r_nv_n),$$

for all  $\epsilon > 0$ , and for all  $f \in \mathcal{F}$ .

$$\text{(C.2)} \quad (r_nv_n)^{-1} \text{Cov}(f(Y_n), g(Y_n)) \longrightarrow c(f, g), \text{ for all } f, g \in \mathcal{F},$$

with  $r_n \ll v_n^{-1} \ll n$ , then the fidis of the empirical process  $(Z_n(f))_{f \in \mathcal{F}}$  of cluster functionals converge to the fidis of a Gaussian process  $(Z(f))_{f \in \mathcal{F}}$  with the covariance function  $c$ .

Drees & Rootzén [2010] have proved CLTs for the process (10). In particular, they have proved CLTs in fidis for this process by using the Bernstein blocks technique together with a  $\beta$ -mixing coupling condition to boil down convergence to convergence of sums over i.i.d. blocks through Eberlein [1984]'s technique involving the metric of total variation. Moreover, Drees & Rootzén [2010] extend the results to the uniform convergence by adding Van Der Vaart & Wellner [1996]'s tightness criteria and asymptotic equicontinuity conditions to the results in fidis that they had obtained.

We aim at extending their CLT's for the empirical process  $(Z_n(f))_{f \in \mathcal{F}}$ , since the family of mixing processes is still very restrictive. One particular and really simple example of a non-mixing process is the AR(1)-process (2). We derive some results as in [Drees & Rootzén, 2010] and some applications as in [Drees, 2011] under much weaker dependence conditions including *eg.* this example.

The  $\tau$ -weak dependence introduced by Dedecker & Prieur [2004a] holds for the Example in eqn. (2), as well as more generally for Bernoulli shifts processes and Markov chains.

### 3. CENTRAL LIMIT THEOREMS FOR CLUSTER FUNCTIONALS

#### 3.1. Fidis convergence

First we give the convergence in fidis. In this case, the technique used to prove the convergence of the empirical process (10) is also the Bernstein blocks technique. In order to do this, we need to extract from each block  $Y_{n,j}$  of length  $r_n$  a sub-block of length  $l_n$ , in such a way that  $l_n = o(r_n)$ . Then we use the remaining sub-blocks, separated by  $l_n$  variables, combined with convenient conditions of  $\tau$ -dependence to couple independent blocks to the original blocks (originally dependent), and thus obtain the CLTs through classic tools.

For this, it is necessary for the triangular array  $(X_{n,i})_{1 \leq i \leq n, n \in \mathbb{N}}$  to satisfy the relation:

$$(D.1) \quad \tau_{1,n}(l_n) = o(r_n^{-1}),$$

such that

$$(B.1) \quad l_n \ll r_n \ll v_n^{-1} \ll n \text{ where } l_n \longrightarrow \infty \text{ and } nv_n \longrightarrow \infty \text{ as } n \rightarrow \infty, \text{ and}$$

$$(B.2) \quad \mathbb{E}(\|X_{n,1}\| \|X_{n,1} \neq 0\|) < \infty.$$

On the other hand, we must not forget the influence of the small blocks extracted with length  $l_n$  over the empirical process (10). In this case, in order to introduce assumptions over these small blocks, it is necessary to consider the following notations, which we will also use throughout the rest of this paper.

**Notation 1** Let  $x = (x_1, x_2, \dots, x_r)$ . We will use the notation  $x^{[l:k]}$  as follows

$$x^{[l:k]} = \begin{cases} 0, & \text{if } r < l, \\ (x_l, \dots, x_k), & \text{if } 1 \leq l \leq k \leq r, \\ Y, & \text{if } k > r. \end{cases}$$

and  $x^{[k]} := x^{[1:k]}$ . Moreover, if  $f \in \mathcal{F}$  is a cluster functional, then we denote

$$\Delta_n(f) := f(Y_n) - f(Y_n^{[r_n - l_n]}), \quad (11)$$

where  $r_n$  is the length of the block  $Y_n$  and  $l_n$  is such that  $l_n = o(r_n)$ .

The following assumption guarantees that the extraction of the small blocks does not disturb the result of the convergence in fidis (if that is the case) of the EPCF  $(Z_n(f))_{f \in \mathcal{F}}$ .

(C.3) For all  $f \in \mathcal{F}$ ,

$$\begin{aligned} \mathbb{E}|\Delta_n(f) - \mathbb{E}\Delta_n(f)|^2 \mathbb{1}\{|\Delta_n(f) - \mathbb{E}\Delta_n(f)| \leq \sqrt{nv_n}\} &= o(r_nv_n) \\ \mathbb{P}\{|\Delta_n(f) - \mathbb{E}\Delta_n(f)| > \sqrt{nv_n}\} &= o(r_n/v_n). \end{aligned}$$

**Theorem 1** Suppose that (B.1)-(B.2), (C.1)-(C.3) and (D.1) hold. Then the fidis of the cluster functionals empirical process  $(Z_n(f))_{f \in \mathcal{F}}$  converge to the fidis of a centered Gaussian process  $(Z(f))_{f \in \mathcal{F}}$  with covariance function  $c$  defined in assumption (C.2).

**Remark 1** Note that Assumptions (C.1) and (C.3) are difficult to check in general, for that reason, consider the following (more restrictive but easier to verify) alternatives conditions:

$$(A.1) \quad \text{Var}(\Delta_n(f)) = o(r_nv_n) \text{ for all } f \in \mathcal{F}.$$

$$(A.3) \quad \mathbb{E}(f(Y_n))^{2+\delta} = O(r_nv_n) \text{ for some } \delta > 0 \text{ and for all } f \in \mathcal{F}.$$

**Lemma 1** The conditions (A.1) and (A.3) implies the conditions (C.1) and (C.3), respectively.

For the proof of this lemma see Lemma 5.2 in [Drees & Rootzén, 2010].

Generally, the (C.2) convergence can be easily verified. However, in some situations it could be difficult to give the limit  $c$  in an explicit mode. In this sense, it is common to use the "tail chain" associated to the original process  $(X_i)_{i \in \mathbb{N}}$ . This terminology was used by Perfekt [1994], but generalized by Yun [1998] to make explicit representations of the extremal index of a higher-order stationary Markov chain. Segers [2003] generalizes this result to stationary sequences with suitable conditions.

We will follow Segers [2003] rationale to give an explicit representation of the covariance function  $c$  through the tail chains. The following proposition (which is a similar result to Segers [2003]'s Theorems 1 and 3) provide some conditions, sufficient to verify (C.2) and which in some situations are easier to prove. The alternative expression to the covariance function  $c$  (defined in (C.2)), is shown below in Corollary 1.

In order to carry this out, it is necessary to consider the following assumption:

**(C.2')** There is a sequence  $W = (W_i)_{i \geq 1}$  of  $E$ -valued random variables such that, for all  $k \in \mathbb{N}$ , the joint conditional distribution

$$P_{(X_{n,i}, \mathbb{1}\{X_{n,i}=0\})_{1 \leq i \leq k} | X_{n,1} \neq 0}$$

converges weakly to  $P_{(W_i, \mathbb{1}\{W_i=0\})}$ , and for all  $f \in \mathcal{F}$  are a.s. continuous with respect to the distribution of  $W^{[k]} = (W_1, \dots, W_k)$  and  $W^{[2:k]} = (W_2, \dots, W_k)$  for all  $k$ , that is,

$$\mathbb{P}\{W^{[2:k]} \in D_{f,k-1}, W_i = 0, \forall i > k\} = \mathbb{P}\{W^{[k]} \in D_{f,k}, W_i = 0, \forall i > k\} = 0$$

where we denote by  $D_{f,k}$  the set of discontinuities of  $f|_{E^k}$ .

**Remark 2** The existence of such sequence  $W$  is guaranteed in particular from Theorem 2 in [Segers, 2003] with  $E = \mathbb{R}$  and the normalization (6). There, Segers has shown that if

$$\mathbb{P}_{((X_{n,i})_{1 \leq i \leq k} | X_1 > u_n)} \xrightarrow{n \rightarrow \infty} -\log G_k,$$

where  $G_k$  is some  $k$ -dimensional extreme value distribution for all  $k \in \mathbb{N}$ , then there exists such "tail chain"  $W = (W_i)_{i \in \mathbb{N}}$  such that

$$P_{((X_{n,i}, \mathbb{1}\{X_{n,i}=0\})_{1 \leq i \leq k} | X_1 > u_n)} \xrightarrow[n \rightarrow \infty]{w} P_{(W_i, \mathbb{1}\{W_i=0\})_{1 \leq i \leq k}}, \quad (12)$$

for all  $k \in \mathbb{N}$ .

**Proposition 1** Suppose that the r.v's  $(X_{n,i})_{1 \leq i \leq n, n \in \mathbb{N}}$  satisfies the following condition:

**(D.2)** There exists  $p, q \geq 1$  with  $p^{-1} + q^{-1} = 1$  and  $\alpha > 0$ , such that

$$\lim_{l \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{r_n \tau_{p,n}(l)}{v_n^{1/p+\alpha}} = 0 \quad \text{and} \quad r_n v_n^{1/q} \xrightarrow[n \rightarrow \infty]{} 0 \quad (13)$$



Then,

$$\mathbb{E}[f(Y_n)|Y_n \neq 0] = \theta_n^{-1} \mathbb{E}\left[f(Y_{n,1}) - f(Y_{n,1}^{[2:r_n]})|X_{n,1} \neq 0\right] + o(1),$$

where  $o(1)$  converges to 0 as  $n \rightarrow \infty$  uniformly for all bounded cluster functionals  $f \in \mathcal{F}$ , and

$$\theta_n := \frac{\mathbb{P}\{Y_n \neq 0\}}{r_n v_n} = \mathbb{P}\{Y_{n,1}^{[2:r_n]} = 0 | X_{n,1} \neq 0\} (1 + o(1)). \quad (14)$$

Additionally, if the assumption (C.2') is satisfied, then:

$$\begin{aligned} m_W &:= \sup\{i \geq 1 : W_i \neq 0\} < \infty, \\ \theta_n &\xrightarrow{n \rightarrow \infty} \theta := \mathbb{P}\{W_i = 0, \forall i \geq 2\} = \mathbb{P}\{m_W = 1\} > 0, \\ P_{f(Y_n)|Y_n \neq 0} &\xrightarrow[n \rightarrow \infty]{w} \frac{1}{\theta} \left( \mathbb{P}\{f(W) \in \cdot\} - \mathbb{P}\{f(W^{[2:\infty]}) \in \cdot, m_W \geq 2\} \right). \end{aligned}$$

**Corollary 1** Suppose that

$$\mathcal{F} = \{f | (f(Y_n))^2\}_{n \in \mathbb{N}} \text{ is uniformly integrable under } P(\cdot)/r_n v_n\}.$$

Assume that (B.1), (B.2), (C.2'), (C.3) and (D.2) hold. Then the fidis of the cluster functionals empirical process  $(Z_n(f))_{f \in \mathcal{F}}$  converge to the fidis of a centered Gaussian process  $(Z(f))_{f \in \mathcal{F}}$  with covariance function  $c$  defined by

$$c(f, g) = \mathbb{E} \left[ (fg)(W) - (fg)(W^{[2:\infty]}) \right]. \quad (15)$$

There are many cases in which  $\|f\|_\infty = \sup_{x \in E_\cup} |f(x)| < \infty$ , for all  $f \in \mathcal{F}$ . Under this condition, it is clear that the conditions (C.1) and (C.3) are satisfied. Therefore, it is important to note the following corollary.

**Corollary 2** Suppose that (B.1), (B.2), (C.2') and (D.2) are satisfied. Then, if  $\|f\|_\infty = \sup_{x \in E_\cup} |f(x)| < \infty$  for all  $f \in \mathcal{F}$ , the fidis of the cluster functionals empirical process  $(Z_n(f))_{f \in \mathcal{F}}$  converges to the fidis of a centered Gaussian process  $(Z(f))_{f \in \mathcal{F}}$  with covariance function  $c$  defined by (15).

## 3.2. Uniform convergence

To prove uniform convergence, we use either asymptotic tightness of  $Z_n$  in the space  $\ell^\infty(\mathcal{F})$ , or asymptotic equicontinuity conditions, by some results in § 2.11 of Van Der Vaart & Wellner [1996]. Those results need independence therefore a argument of coupling for the blocks  $(Y_{n,j})_{1 \leq j \leq m_n, n \in \mathbb{N}}$  is also used here.

### 3.2.1 Asymptotic tightness

**Definition 1** The sequence  $(Z_n)_{n \in \mathbb{N}}$  is asymptotically tight if for every  $\epsilon > 0$  there exists a compact set  $K \in \ell^\infty(\mathcal{F})$  such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}^*(Z_n \notin K^\delta) < \epsilon, \quad \text{for every } \delta > 0,$$

where  $K^\delta = \{f \in \ell^\infty(\mathcal{F}) : d_{\mathcal{F}}(f, K) < \delta\}$  is the " $\delta$ -enlargement" around  $K$  and  $\mathbb{P}^*$  denotes the outer probability.

**Definition 2** The bracketing number  $N_{[\cdot]}(\epsilon, \mathcal{F}, L_2^n)$  is defined as the smallest number  $N_\epsilon$  such that for each  $n \in \mathbb{N}$  there exists a partition  $(\mathcal{F}_{n,k}^\epsilon)_{1 \leq k \leq N_\epsilon}$  of  $\mathcal{F}$  such that

$$\mathbb{E}^* \sup_{f,g \in \mathcal{F}_{n,k}^\epsilon} (f(Y_n) - g(Y_n))^2 \leq \epsilon^2 r_n v_n, \quad \text{for } 1 \leq k \leq N_\epsilon,$$

where  $\mathbb{E}^*$  denotes the outer expectation.

In order to use Theorem 2.11.9 in [Van Der Vaart & Wellner, 1996] we need:

**(T.1)** The set  $\mathcal{F}$  of cluster functionals is such that for each  $f \in \mathcal{F}$  the expression  $\mathbb{E}f^2(Y_n)$  is finite for all  $n \in \mathbb{N}$  and such that the envelope function satisfies:

$$F(x) := \sup_{f \in \mathcal{F}} |f(x)| < \infty, \quad \forall x \in E_\cup.$$

**(T.2)**  $\mathbb{E}^*(F(Y_n)\mathbb{1}\{F(Y_n) > \epsilon\sqrt{n v_n}\}) = o(r_n\sqrt{v_n/n})$  for all  $\epsilon > 0$ .

Note that for a sequence of monotonically increasing positive functions  $(h_n(\delta))_{n \geq 1}$  the convergence of  $h_n(\delta_n)$  to zero  $\forall \delta_n \downarrow 0$  is equivalent to

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} h_n(\delta) = 0,$$

thus the Assumptions 2 and 3 of Theorem 2.11.9 from [Van Der Vaart & Wellner, 1996] are reformulated as follows:

**(T.3)** There exists a semi-metric  $\rho$  on  $\mathcal{F}$  such that  $\mathcal{F}$  is totally bounded with respect to (w.r.t.)  $\rho$  and

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{f,g \in \mathcal{F}: \rho(f,g) < \delta} \frac{1}{r_n v_n} \mathbb{E} (f(Y_n) - g(Y_n))^2 = 0.$$

**(T.4)**

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \int_0^\delta \sqrt{\log N_{[\cdot]}(\epsilon, \mathcal{F}, L_2^n)} d\epsilon = 0.$$

**Theorem 2** Suppose that (B.1), (B.2), (D.1) hold and that (T.1) - (T.4) are satisfied. Then the empirical process  $(Z_n)_{n \in \mathbb{N}}$  is asymptotically tight in  $\ell^\infty(\mathcal{F})$ . Moreover, if the assumptions (C.1)-(C.3) hold, then  $Z_n$  converges to a centered Gaussian process  $Z$  with covariance function  $c$  in (C.2).

### 3.2.2 Asymptotic equicontinuity

**Definition 3** The sequence  $(Z_n)_{n \in \mathbb{N}}$  is asymptotically equicontinuous w.r.t. a semi-metric  $\rho$  if for any  $\epsilon > 0$  and  $\eta > 0$  there exists some  $\delta > 0$  such that:

$$\limsup_{n \rightarrow \infty} \mathbb{P}^* \left( \sup_{f,g \in \mathcal{F}: \rho(f,g) < \delta} |Z_n(f) - Z_n(g)| > \epsilon \right) < \eta.$$

We use Theorem 2.11.1 in [Van Der Vaart & Wellner, 1996] to prove asymptotic equicontinuity. In order to do this, we need to define a semi-metric  $\rho_n$  on  $\mathcal{F}$  as follows. Let  $(Y_{n,j}^*)_{1 \leq j \leq m_n}$  be the independent copies of  $(Y_{n,j})_{1 \leq j \leq m_n}$ . We define  $\rho_n$  as:

$$\rho_n(f, g) := \sqrt{\frac{1}{nv_n} \sum_{j=1}^{m_n} (f(Y_{n,j}^*) - g(Y_{n,j}^*))^2}. \quad (16)$$

Moreover, we denote by  $N(\epsilon, \mathcal{F}, \rho)$  the "covering number", the minimum number of balls (w.r.t. the semi-metric  $\rho$ ) with radius  $\epsilon > 0$  necessary to cover  $\mathcal{F}$ . In this way, we can add to the list of assumptions the following two:

(T.4') For  $k = 1, 2$  the map

$$(Y_{n,1}^*, \dots, Y_{n,[m_n/2]}^*) \mapsto \sup_{f, g \in \mathcal{F}: \rho(f, g) < \delta} \sum_{j=1}^{[m_n/2]} e_j \left( f(Y_{n,j}^*) - g(Y_{n,j}^*) \right)^k$$

is measurable for each  $\delta > 0$ , each vector  $(e_1, \dots, e_{[m_n/2]}) \in \{-1, 0, 1\}^{[m_n/2]}$  and each  $n \in \mathbb{N}$ .

(T.5)

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}^* \left( \int_0^\delta \sqrt{\log N(\epsilon, \mathcal{F}, \rho_n)} d\epsilon > \eta \right) = 0, \quad \forall \eta > 0.$$

**Theorem 3** Suppose that (B.1), (B.2), (D.1) hold and that (T.1)-(T.3), (T.4') and (T.5) are satisfied. Then the empirical process  $(Z_n)_{n \in \mathbb{N}}$  is asymptotically equicontinuous. Moreover if the assumptions (C.1)-(C.3) hold, then  $Z_n$  converges to a centered Gaussian process  $Z$  with covariance function  $c$  in (C.2).

### 3.3. Application: blocks estimator of the extremal index

Let  $(X_i)_{i \in \mathbb{N}}$  be a real stationary time series with distribution function  $F$ . Now consider the index defined in (14) with the extreme normalization (6) and  $u_n := F^{\leftarrow}(1 - v_n t)$ , for all  $t \in [0, 1]$ , i.e.

$$\theta_{n,t} := \frac{\mathbb{P}\{Y_n \neq 0\}}{r_n v_n t} = \frac{\mathbb{P}\{\max_{1 \leq i \leq r_n} X_i > u_n\}}{r_n v_n t}, \quad \text{with } t \in (0, 1]. \quad (17)$$

Note that if  $r_n$  satisfies (B.1) condition and if (D.2) holds, then by using Proposition 1, there exists a number (extremal index)  $\theta \in (0, 1]$  such that

$$\theta_{n,t} \xrightarrow[n \rightarrow \infty]{} \theta \quad \text{uniformly for } t \in (0, 1]. \quad (18)$$

Given the convergence (18), Drees has suggested in his paper Drees [2011] to estimate  $\theta$  replacing the unknown probability  $\mathbb{P}\{\max_{1 \leq i \leq r_n} X_i > u_n\}$  and the unknown expectation

$r_n v_n t = \mathbb{E} [\sum_{i=1}^{r_n} \mathbb{1}\{X_i > u_n\}]$  by a empirical expression for  $\theta_{n,t}$ :

$$\hat{\theta}_{n,t} := \frac{\sum_{j=1}^{m_n} \mathbb{1}\{\max_{(j-1)r_n < i \leq jr_n} X_i > u_n\}}{\sum_{j=1}^{m_n} \sum_{i=(j-1)r_n+1}^{jr_n} \mathbb{1}\{X_i > u_n\}}, \quad (19)$$

where  $m_n = \lfloor n/r_n \rfloor$  such that  $1 \ll r_n \ll v_n^{-1} \ll n$  but  $nv_n \rightarrow \infty$ . Thus, such estimator (19) (called *blocks estimator of the extremal index*) can be expressed in terms of two empirical processes of cluster functionals  $(Z_n(f_t), Z_n(g_t))_{0 \leq t \leq 1}$  defined in (10). For this, suppose without loss of generality that the random variables  $(X_i)_{1 \leq i \leq n}$  are uniformly distributed on  $[0, 1]$  (otherwise, just consider the transformation  $U_i = F(X_i)$ ,  $1 \leq i \leq n$ , where  $F$  is the distribution function of  $X_1$ , see Drees [2011]). Then, with the normalization (6) such that  $a_n = v_n = 1 - u_n$  and the blocks  $(Y_{n,j})_{1 \leq j \leq m_n}$  defined in (9), we have that

$$\hat{\theta}_{n,t} = \frac{m_n^{-1} \sum_{j=1}^{m_n} f_t(Y_{n,j})}{m_n^{-1} \sum_{j=1}^{m_n} g_t(Y_{n,j})} = \frac{\mathbb{E} f_t(Y_{n,1}) + (nv_n)^{1/2} m_n^{-1} Z_n(f_t)}{\mathbb{E} g_t(Y_{n,1}) + (nv_n)^{1/2} m_n^{-1} Z_n(g_t)}, \quad (20)$$

where

$$f_t(x_1, \dots, x_r) := \mathbb{1}\{\max_{1 \leq i \leq r} x_i > 1 - t\} \quad (21)$$

$$g_t(x_1, \dots, x_r) := \sum_{i=1}^r \mathbb{1}\{x_i > 1 - t\}. \quad (22)$$

For this particular case, we consider the following assumptions:

**(C.2.1)**  $(r_n v_n)^{-1} \text{Cov}(g_s(Y_n), g_t(Y_n)) \rightarrow c_g(s, t)$ , for all  $0 \leq s, t \leq 1$ .

**(C.2.2)**  $(r_n v_n)^{-1} \text{Cov}(f_s(Y_n), g_t(Y_n)) \rightarrow c_{fg}(s, t)$ , for all  $0 \leq s, t \leq 1$ .

**(T)** For some bounded function  $h : (0, 1] \rightarrow \mathbb{R}$  such that  $\lim_{t \rightarrow 0} h(t) = 0$

$$(r_n v_n)^{-1} \mathbb{E} (f_s(Y_{n,1}) - f_t(Y_{n,1}))^2 \leq h(t - s), \quad \forall 0 \leq s < t \leq 1,$$

for all  $n$  sufficiently large.

The following are a slight variation of the first two results of Drees [2011], in the sense that we replace the  $\beta$ -mixing condition with  $\tau$ -dependence condition.

## Proposition 2

(3.1) Suppose that (B.1), (B.2) and (D.1) are satisfied. Then  $(Z_n(f_t))_{0 \leq t \leq 1}$  converges weakly to  $Z_f := (\sqrt{\theta} B_t)_{0 \leq t \leq 1}$ , where  $B$  denote a standard Brownian motion.

(3.2) If additionally (C.2.1) and (T) are satisfied and  $r_n = o(\sqrt{nv_n})$ , then the sequence of processes  $(Z_n(g_t))_{0 \leq t \leq 1}$  converges weakly to a centered Gaussian process  $(Z_g(t))_{0 \leq t \leq 1}$  with covariance function  $c_g$ .

(3.3) Under all the hypothesis of (3.1) and (3.2), if moreover (C.2.2) holds, then  $(Z_n(f_t), Z_n(g_t))_{0 \leq t \leq 1}$  converge weakly to  $(Z_f(t), Z_g(t))_{0 \leq t \leq 1}$  with

$$\begin{aligned} \text{Cov}(Z_f(s), Z_f(t)) &= \theta(s \wedge t), \\ \text{Cov}(Z_g(s), Z_g(t)) &= c_g(s, t), \\ \text{Cov}(Z_f(s), Z_g(t)) &= c_{fg}(s, t), \quad 0 \leq s, t \leq 1. \end{aligned} \quad (23)$$

Using the same argument in Remark 2, we can find explicit expressions for the covariance functions  $c_g$  and  $c_{fg}$  as functions of the "tail chains" of  $(X_i)_{i \in \mathbb{N}}$ . This is, if for every  $k \in \mathbb{N}$  the distribution function of  $(X_1, \dots, X_k)$  belongs to the domain of attraction of an extreme-value distribution, then there exist a sequence  $W = (W_i)_{i \in \mathbb{N}}$  such that (12) hold. In such case:

$$\begin{aligned} c_g(s, t) &= s \wedge t + \sum_{i=1}^{\infty} (\mathbb{P}\{W_1 > 1-s, W_{i+1} > 1-t\} + \mathbb{P}\{W_1 > 1-t, W_{i+1} > 1-s\}) \\ c_{fg}(s, t) &= \begin{cases} \mathbb{P}\{W_1 > 1-t, \max_{j \geq 1} W_j > 1-s\} \\ + \sum_{i=1}^{\infty} \mathbb{P}\{W_1 > 1-s, W_{i+1} > 1-t, \max_{j \geq 2} W_j \leq 1-s\}, & s < t, \\ t & s \geq t. \end{cases} \end{aligned}$$

**Corollary 3** Under Proposition 2 - (3.3)'s assumptions,

$$(\sqrt{nv_n t}(\hat{\theta}_{n,t} - \theta_{n,t}))_{0 < t \leq 1} \xrightarrow[n \rightarrow \infty]{w} Z := Z_f - \theta Z_g, \quad (24)$$

where  $Z$  is a Gaussian process such that  $\mathbb{E}Z(t) = 0$  and

$$\text{Cov}(Z(s), Z(t)) = \theta(s \wedge t - c_{fg}(s, t) - c_{fg}(t, s) + \theta^2 c_g(s, t)). \quad (25)$$

## 4. EXAMPLES AND SIMULATIONS

### 4.1. AR(1)-process with the functional "number of excesses over $x$ "

We consider the AR(1)-process (2) where  $b \geq 2$  is an integer,  $(\xi_i)_{i \in \mathbb{N}}$  are i.i.d. and uniformly distributed on the set  $U(b) := \{0, 1, \dots, b-1\}$ .

It is clear that  $X_0$  is uniformly distributed on  $[0, 1]$ . Moreover we define the normalized random variables  $(X_{n,i})_{1 \leq i \leq n, n \in \mathbb{N}}$  as in eqn. (8) with  $a_n = v_n = 1 - u_n$ . We set  $(x_1, \dots, x_d) \leq (y_1, \dots, y_d)$  if and only if  $x_i \leq y_i$ , for all  $i = 1, \dots, d$  in case  $x, y \in [0, 1]^d$ . Then

$$\begin{aligned} &\mathbb{P}\{X_{n,1} > x | X_{n,1} \neq 0\} \\ &= \frac{1}{b^d \bar{v}_n} \sum_{j_1, \dots, j_d \in U(b)} \left( \max_{i=1, \dots, d} \left\{ 1 - b^i + \sum_{s=1}^i b^{s-1} j_s + b^i v_n (1 - x_i) \right\} \wedge 1 \right) \\ &\xrightarrow[n \rightarrow \infty]{} \max_{i=1, \dots, d} \{b^{i-d} (1 - x_i)\}, \end{aligned} \quad (26)$$

where  $\bar{v}_n := \mathbb{P}\{X_{n,1} \neq 0\} \sim v_n = \mathbb{P}\{X_1 > u_n\} \xrightarrow{n \rightarrow \infty} 0$ .

Consider  $\mathcal{F}$  the family of cluster functionals:

$$\mathcal{F} = \left\{ f_x, x \in [0, 1]^d \right\}, \quad \text{with} \quad f_x(x_1, \dots, x_r) = \sum_{i=1}^r \mathbb{1}\{x_i > x\} \quad (27)$$

For this case, we obtain the covariance function  $c$  of (C.2):

$$\begin{aligned} c(x, y) = & \min \left( \max_{k=1, \dots, d} \{b^k(1 - x_k)\}, \max_{k=1, \dots, d} \{b^k(1 - y_k)\} \right) \\ & + \sum_{i=1}^{\infty} H_{b,i}(x, y) + \sum_{i=1}^{\infty} H_{b,i}(y, x), \end{aligned} \quad (28)$$

where, for  $i \geq d$

$$H_{b,i}(x, y) := \frac{1}{b^i} \min \left( \max_{k=1, \dots, d} \{b^k(1 - x_k)\}, \max_{k=1, \dots, d} \{b^{k+i}(1 - y_k)\} \right) \quad (29)$$

and for  $1 \leq i < d$ ,

$$\begin{aligned} & H_{b,i}(x, y) \\ & := \frac{1}{b^i} \min \left( \max_{k=1, \dots, i} \{b^k(1 - x_k)\}, \max_{k=i+1, \dots, d} \{b^k \min(1 - x_k, 1 - y_k)\}, \max_{k=d-i, \dots, d} \{b^{k+i}(1 - y_k)\} \right) \end{aligned}$$

Conditions (C.1), (C.3), (T.1)-(T.4) hold for uniformly distributed random variables and for the same family  $\mathcal{F}$ , see page 2177 and 2178 in Drees & Rootzén [2010]. Thus, under assumption (B.1), setting  $l_n$  and  $r_n$  such that  $b^{-l_n} = o(r_n^{-1})$  (see Appendix A.1 - Application 1) then the empirical process  $(Z_n(x))_{x \in [0, 1]^d}$  defined as:

$$\begin{aligned} Z_n(x) &:= \frac{1}{\sqrt{nv_n}} \sum_{i=1}^{r_n m_n} (\mathbb{1}\{X_{n,i} > x\} - \mathbb{P}\{X_{n,i} > x\}) \\ &\sim \frac{1}{\sqrt{nv_n}} \sum_{i=1}^n (\mathbb{1}\{X_{n,i} > x\} - \mathbb{P}\{X_{n,i} > x\}) \end{aligned} \quad (30)$$

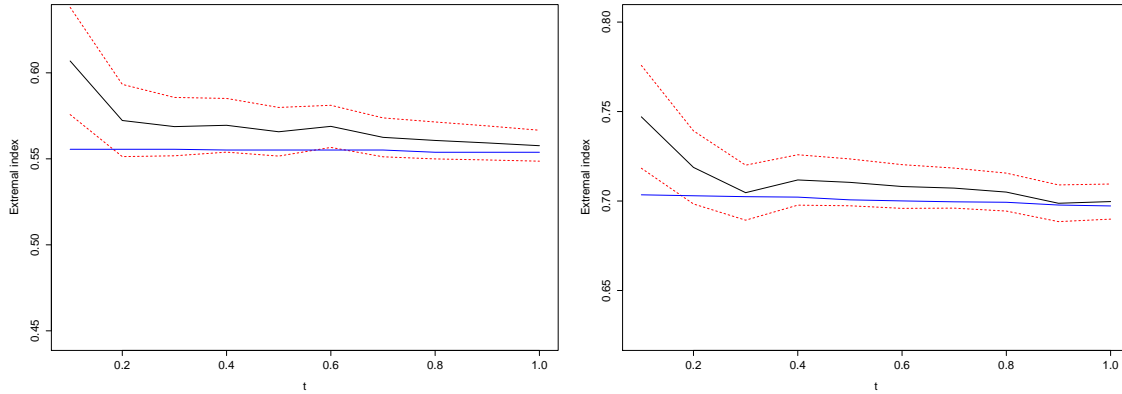
converges to a centered Gaussian process  $Z$  with covariance function (28).

## 4.2. Simulation study

The experiment is to estimate the extremal index  $\theta$  through the blocks estimator of the extremal index (20).

Let us consider the AR(1)-process (2). Here, as  $X_0$  is uniformly distributed on  $[0, 1]$  and  $X_i = \frac{X_0}{b^i} + \sum_{s=1}^i \frac{\xi_s}{b^{i-s+1}}$  for all  $i \geq 1$ , we will take this to obtain a theoretical expression for the index (17) with  $u_n = 1 - v_n t$  for  $t \in (0, 1]$ :

$$\theta_{n,t} = \frac{1}{b^{r_n} r_n v_n t} \sum_{j_1, \dots, j_{r_n} \in U(b)} \min \left( \max_{i=1, \dots, r_n} \left\{ 1 - b^i(1 - v_n t) + \sum_{s=1}^i b^{s-1} j_s \right\}_+, 1 \right), \quad (31)$$



**Figure 1:** *Extremal index for the AR(1)-input*

Left:  $\theta_{n=10^4,t}$  is the blue curve,  $\hat{\theta}_{n=10^4,t}$  is the black curve and the confidence intervals  $I_t = (CI_i(t), CI_s(t))$  are the red curves, for the AR(1)-input (2) with  $b = 2$ . Right: the same situation but with  $b = 3$

which converges to some  $\theta = \theta(b) \in (0, 1)$  if (B.1) is satisfied with  $b^{-l_n} = o(r_n^{-1}v_n^\beta)$ , for some  $\beta > 0$ .

We will use the advantage of having this theoretical expression (31) of the extremal index  $\hat{\theta}_{n,t}$  to compare it to its asymptotic estimations  $\hat{\theta}_{n,t}$ . For this, we simulate AR(1)-processes (2) for  $b = 2, 3$  and their blocks estimators (20) respective with the normalization (6) taking  $a_n = v_n = 1 - u_n$  to make the comparison of the results estimated with the theoretic model (31).

Let us suppose we have data of a size  $M = 600000$  which adjusts appropriately to an AR(1)-process (2) for  $b = 2$  ( $b = 3$ ). Because the process is stationary, we can divide this data into  $N = 60$  blocks of length  $n = 10^4$ . Moreover, we choose a threshold  $u_n$  such that  $v_n = n^{-1/2}$  and the sub-blocks of length  $r_n = \lfloor \log(n) \rfloor$ . In Figure 1 we showed a polygonal curve  $(t, \theta_{n=10^4,t})_{t=0.1,0.2,\dots,1}$  (blue curve) of  $(t, \theta_{n=10^4,t})_{0 \leq t \leq 1}$  and a mean polygonal curve  $(t, \hat{\theta}_{n=10^4,t})_{t=0.1,0.2,\dots,1}$  (black curve) of  $(t, \hat{\theta}_{n=10^4,t})_{0 \leq t \leq 1}$ . Note that for  $n = 10^4$ , the symmetry of the confidence band  $(CI_i(t), CI_s(t))_{0 \leq t \leq 1}$  with respect to  $(\hat{\theta}_{n,t})_{0 \leq t \leq 1}$  and with a confidence level  $1 - \alpha = 0.95$  (red curves), already shows the gaussian behaviour of the estimator. Furthermore, as expected, the estimated value through the blocks estimator is quite close to the extremal index theoretical (31), with  $n = 10^4$ . The numerical results are shown in Tables 1 and 2, for the cases  $b = 2$  and  $b = 3$ , respectively.

t	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\theta_{n,t}$	0.555	0.555	0.555	0.555	0.555	0.555	0.555	0.553	0.553	0.553
$\widehat{\theta}_{n,t}$	0.606	0.572	0.568	0.569	0.565	0.568	0.562	0.560	0.559	0.557
$CI_s$	0.638	0.593	0.585	0.585	0.579	0.581	0.573	0.571	0.569	0.566
$CI_i$	0.575	0.551	0.551	0.553	0.551	0.556	0.551	0.549	0.549	0.548

**Table 1:** Comparison between the blocks estimation and the theoretical approximation (31), for the AR(1)-input with  $b = 2$  and  $n = 10^4$ .

t	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\theta_{n,t}$	0.703	0.702	0.702	0.702	0.700	0.700	0.699	0.699	0.697	0.697
$\widehat{\theta}_{n,t}$	0.747	0.718	0.704	0.711	0.710	0.708	0.707	0.704	0.698	0.699
$CI_s$	0.775	0.739	0.720	0.725	0.723	0.720	0.718	0.715	0.708	0.709
$CI_i$	0.718	0.698	0.689	0.697	0.697	0.695	0.696	0.694	0.688	0.689

**Table 2:** Comparison between the blocks estimation and the theoretical approximation (31), for the AR(1)-input with  $b = 3$  and  $n = 10^4$ .

## APPENDIX. $\tau$ -WEAK DEPENDENCE AND PROOFS

### A.1. Brief interlude into $\tau$ -weak dependence

Let  $(\Omega, \mathcal{A}, \mathbb{P})$  be a probability space, and  $\mathcal{M}$  a  $\sigma$ -algebra of  $\mathcal{A}$ . Let  $(E, \delta)$  be a Polish space endowed with its metric. For any  $E$ -valued random variable  $X$ ,  $\mathbb{L}^p$ -integrable (*i.e.*  $X$  satisfies  $\|\delta(X, 0)\|_p := (\int \delta^p(x, 0) \mathbb{P}(dx))^{1/p} < \infty$ ), Dedecker & Prieur [2004a] defined the **coefficient**  $\tau_p$  as:

$$\tau_p(\mathcal{M}, X) := \|\sup \{ \mathbb{E}[h(X)|\mathcal{M}] - \mathbb{E}[h(X)] : h \in \Lambda(E, \delta) \} \|_p \quad (32)$$

where  $\Lambda(E, \delta)$  denotes the class of all Lipschitz functions  $h : E \rightarrow \mathbb{R}$  such that

$$\text{Lip}(h) := \sup_{x \neq y} \frac{|h(x) - h(y)|}{\delta(x, y)} \leq 1.$$

Let  $\mathcal{X} = (X_{n,i})_{1 \leq i \leq n, n \in \mathbb{N}}$  be a triangular array of  $\mathbb{L}^p$ -integrable  $E$ -valued random variables, and  $(\mathcal{M}_i)_{i \in \mathbb{Z}}$  be a sequence of  $\sigma$ -algebras of  $\mathcal{A}$ .

Then, for any  $n \in \mathbb{N}$ , we define the coefficient:

$$\tau_{p,n}(k) := \sup_{l \geq 1} l^{-1} \sup \{ \tau_p(\mathcal{M}_i, (X_{n,j_1}, \dots, X_{n,j_l})) : i + k \leq j_1 < \dots < j_l \leq n \}, \quad (33)$$

where we consider the distance

$$\delta_l(x, y) = \sum_{i=1}^l \delta(x_i, y_i) \quad (34)$$

on  $E^l$ . Moreover, we say that  $\mathcal{X}$  is  $\tau_p$ -weakly dependent if

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \tau_{p,n}(k) = 0. \quad (35)$$



**Remark 3** Recall that  $\mathcal{X} = (X_{n,i})_{1 \leq i \leq n, n \in \mathbb{N}}$  is constructed from a random process  $\mathbb{X} = (X_i)_{i \in \mathbb{Z}}$ . Therefore the dependence properties of  $\mathcal{X}$ 's are inherited from those of  $\mathbb{X}$ . Even more so, if  $\mathbb{X}$  is  $\tau_p$ -weakly dependent (in the usual sense defined by Dedecker & Prieur [2004a] for random processes), then  $\mathcal{X}$  is  $\tau_p$ -weakly dependent with  $\tau_{p,n}(\cdot) = L_n \cdot \tau_p(\cdot)$ , for some positive constant  $L_n$  (which is written in function of  $\mathcal{X}$ 's normalization constants). For instance, if we consider the normalizations in eqn. (6)-(8) we obtain that  $\tau_{p,n}(\cdot) \leq a_n^{-1} \cdot \tau_p(\cdot)$ .

In this sense, if we want to study  $\mathcal{X}$ 's  $\tau$ -dependence properties, suffice it to take into account  $\mathbb{X}$ 's  $\tau$ -dependence properties.

We make use of the previous remark to mention the following examples of  $\tau_p$ -weakly dependent processes without considering the normalizations.

**Example 1 (Causal Bernoulli shifts)** Let  $(\xi_i)_{i \in \mathbb{Z}}$  be a sequence of i.i.d.r.v's. (independent and identically distributed random variables) with values in a measurable space  $D$ . Assume that there exists a function  $H : D^{\mathbb{N}} \rightarrow \mathbb{R}$ , such that  $H(\xi_0, \xi_{-1}, \dots)$  is defined almost surely. Then the stationary sequence  $(X_i)_{i \geq 0}$  defined by  $X_i = H(\xi_i, \xi_{i-1}, \dots)$  is called a causal Bernoulli shifts.

Let  $(\xi'_i)_{i \in \mathbb{Z}}$  be an independent copy of the i.i.d. sequence  $(\xi_i)_{i \in \mathbb{Z}}$ . Consider a decreasing sequence  $(\Delta_p(i))_{i \geq 0}$  such that

$$\|X_i - X'_i\|_p \leq \Delta_p(i) \quad (36)$$

for some  $p \in [1, \infty]$ , where  $X_i = H(\xi_i, \xi_{i-1}, \dots)$  and  $X'_i = H(\xi_i, \dots, \xi_1, \xi'_0, \xi'_1, \dots)$ . Then, if  $\mathcal{M}_i = \sigma(X_j : j \leq i)$ , the coefficient  $\tau_p(k)$  of  $(X_i)_{i \in \mathbb{Z}}$  is bounded above by  $\Delta_p(k)$ , for all  $k \in \mathbb{N}$ .

**Application 1 (Causal linear processes)** Let  $D = \mathbb{R}$  and

$$X_i = \sum_{j=0}^{\infty} b_j \xi_{i-j}. \quad (37)$$

Here we set  $\Delta'_p(i) = 2\|\xi_0\|_p \sum_{j \geq i} |b_j| \geq \Delta_p(i)$  in case  $\|\xi_0\|_p < \infty$ .

The model (2) can be write as (37) with  $b_j = b^{-j-1}$  for some integer  $b \geq 2$ . Since  $\xi_0$  is uniformly distributed on  $U(b) = \{0, \dots, b-1\}$  in this case, then  $X_0$  is uniformly distributed over  $[0, 1]$  and  $\Delta_{\infty}(i) \leq b^{-i}$ .

**Example 2 (Markov models)** Let  $G : (\mathbb{R}^l, \mathcal{B}(\mathbb{R}^l)) \times (D, \mathcal{D}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  be a measurable function and let  $(X_i)_{i \geq 1-l}$  be a sequence of random variables with values in  $\mathbb{R}$  such that

$$X_i = G(X_{i-1}, X_{i-2}, \dots, X_{i-l}; \xi_i), \quad \forall i \geq 1, \quad (38)$$

for some sequence  $(\xi_i)_{i \in \mathbb{N}}$  of i.i.d.r.v's. with values in a measurable space  $D$  and independent of  $(X_0, \dots, X_{1-l})$ . Then the random variables  $Y_i = (X_i, X_{i-1}, \dots, X_{i-l+1})$  defines a Markov chain such that  $Y_i = F(Y_{i-1}; \xi_i)$  with

$$F(x_l, \dots, x_1; \xi) := (G(x_l, \dots, x_1; \xi), x_l, x_{l-1}, \dots, x_2). \quad (39)$$

Assume that  $Y_0 = (X_0, \dots, X_{1-l})$  is a stationary solution of (38). Let  $Y'_0 = (X'_0, \dots, X'_{1-l})$  be independent of  $(Y_0, (\xi_i)_{i \in \mathbb{N}})$  and distributed as  $Y_0$ . Then setting

$$X'_i = G(X'_{i-1}, \dots, X'_{i-l}; \xi_i), \quad (40)$$

$X'_i$  is distributed as  $X_i$  and it is independent of  $\mathcal{M}_0 = \sigma(X_0, \dots, X_{1-l})$ , for all  $i \in \mathbb{N}$ . As in the previous example, let  $(\Delta_p(i))_{i \geq 0}$  be a non increasing sequence such that (36) holds, with  $X_i$  and  $X'_i$  defined in (38) and (40), respectively. Hence one can apply the result of Lemma 3 in [Dedecker & Prieur, 2004a], and we obtain that  $\tau_p(k) \leq \Delta_p(k)$ .

In particular if  $G$  is such that

$$\|G(x; \xi_1) - G(y; \xi_1)\|_p \leq \sum_{i=1}^l a_i |x_i - y_i|, \quad \text{with } \sum_{i=1}^l a_i < 1, \quad (41)$$

then  $\Delta_p(k) \leq Ca^k$  for some  $a \in [0, 1)$  and some  $C > 0$ . (see [Dedecker et al., 2007], page 34).

**Application 2 (Contracting Markov chain)** Let  $X_i = G(X_{i-1}, \xi_i)$  be a Markov chain such that  $G : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \times (D, \mathcal{D}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  is a measurable function and

$$A = \|G(0; \xi_1)\|_p < \infty \text{ and } \|G(x; \xi_1) - G(y; \xi_1)\|_p \leq a|x - y|, \quad (42)$$

for some  $a \in (0, 1)$  and some  $p \in [1, \infty]$ . Then,  $(X_i)_{i \in \mathbb{N}}$  has a stationary solution with  $p$ -th order finite moment as this is proved on page 35 of Dedecker et al. [2007]. Moreover under this condition:  $\Delta_p(k) = \|X'_0 - X_0\|_p \cdot a^k$ .

**Remark 4** In particular if  $G(u; \xi) = A(u) + B(u)\xi$  for suitable Lipschitz functions  $A(u)$  and  $B(u)$  with  $u \in \mathbb{R}$ , then the corresponding iterative model (ARCH-type process)  $X_i = G(X_{i-1}; \xi_i)$  satisfies (42) with  $a = \text{Lip}(A) + \|\xi_1\|_p \text{Lip}(B) < 1$ .

**Remark 5** The stationary iterative models  $X_i = G(X_{i-1}, \xi_i)$  are causal Bernoulli shifts if the condition (42) holds; this is proved in Proposition 3.2 in [Dedecker et al., 2007].

**Application 3 (Nonlinear AR( $l$ )-models)** Let  $l \geq 1$  and  $(X_i)_i$  be the stationary solution of some equation

$$X_i = R(X_{i-1}, \dots, X_{i-l}) + \xi_i$$

for some measurable function  $R : \mathbb{R}^l \rightarrow \mathbb{R}$ . The process  $(X_i)_i$  is then called a stationary real nonlinear autoregressive model of order  $l$ . If  $\|\xi_1\|_p < \infty$  and

$$|R(u_1, \dots, u_l) - R(v_1, \dots, v_l)| \leq \sum_{i=1}^l a_i |u_i - v_i|, \text{ for } a_1, \dots, a_l \geq 0 \text{ with } \sum_{i=1}^l a_i < 1,$$

and for all  $(u_1, \dots, u_l), (v_1, \dots, v_l) \in \mathbb{R}^l$ , then the function  $G : \mathbb{R}^{l+1} \rightarrow \mathbb{R}$  defined by  $G(u; \xi) = R(u) + \xi$  satisfies Condition (41) and therefore the sequence  $(\Delta_p(k))_k$  admits an exponential decay rate.

## A.2. Coupling

Let  $d_{\cup} : E_{\cup} \times E_{\cup} \longrightarrow [0, \infty)$  be a pseudometric defined as follows: let  $x_k = (x_k^{(1)}, x_k^{(2)}, \dots, x_k^{(r_k)}) \in E_{\cup}$  for  $k = 1, 2$  and we denote

$$x_{1,2} = \begin{cases} x_1, & \text{if } r_1 \geq r_2, \\ (x_1^{(1)}, \dots, x_1^{(r_1)}, x_2^{(r_1+1)}, \dots, x_2^{(r_2)}), & \text{if } r_1 < r_2. \end{cases}$$

Similarly is denoted  $x_{2,1}$ . Then,

$$d_{\cup}(x_1, x_2) := d_{r_1 \vee r_2}(x_{1,2}, x_{2,1}), \quad (43)$$

where  $d_r(\cdot, \cdot)$  is defined in (34).

**Lemma 2 (Coupling: the even and odd blocks)** *Suppose that the random variables  $(X_{n,i})_{1 \leq i \leq n}$  are such that (B.2) holds. We consider together even and odd block sizes, by using  $k = 0$  or  $1$  according to the parity. Assume that for each  $j \in \{2, \dots, \lfloor m_n/2 \rfloor\}$  there is a random variable  $U_{k,j}$  uniformly distributed on  $[0, 1]$  and independent of  $\mathcal{M}_{n,j-1}^k = \sigma(Y_{n,2-k}, \dots, Y_{n,2(j-1)-k})$  and  $\sigma(Y_{n,2j-k})$ . Then there exists a random block  $\dot{Y}_{n,2j-k}$  measurable with respect to  $\mathcal{M}_{n,j-1}^k \vee \sigma(Y_{n,2j-k}) \vee \sigma(U_{k,j})$ , independent of  $\mathcal{M}_{n,j-1}^k$  and distributed as  $Y_{n,2j-k}$  such that*

$$\left\| \mathbb{E}(d_{\cup}(Y_{n,2j-k}, \dot{Y}_{n,2j-k})) \mid \mathcal{M}_{n,j-1}^k \right\|_1 \leq r_n \tau_{1,n}(r_n). \quad (44)$$

*In particular, if we set  $\mathcal{M}_{n,j-1}^k = \sigma(\dot{Y}_{n,2-k}, \dots, \dot{Y}_{n,2(j-1)-k})$ , then the blocks  $(\dot{Y}_{n,2j-k})_{1 \leq j \leq \lfloor m_n/2 \rfloor}$  are independent.*

**Proof:** We set here  $k = 0$  (for even block sizes) since the steps are similar if  $k = 1$ . Let  $Y_{n,2j} = (X_{n,i})_{i \in A_j}$  be a random block, where  $A_j = \{(2j-1)r_n + 1, \dots, 2jr_n\}$ . Then, from Dedecker et al. [2007]'s Lemma 5.3, there exists a random random block  $\dot{Y}_{n,2j} \in E^{r_n}$  measurable with respect to  $\mathcal{M}_{n,j-1}^0 \vee \sigma(Y_{n,2j}) \vee \sigma(U_{0,j})$  independent of  $\mathcal{M}_{n,j-1}^0$  and distributed as  $Y_{n,2j}$  such that

$$\tau_p(\mathcal{M}_{n,j-1}^0, Y_{n,2j}) = \left\| \mathbb{E}(d_{r_n}(Y_{n,2j}, \dot{Y}_{n,2j}) \mid \mathcal{M}_{n,j-1}^0) \right\|_p. \quad (45)$$

Therefore,

$$\begin{aligned} \left\| \mathbb{E}(d_{\cup}(Y_{n,2j}, \dot{Y}_{n,2j}) \mid \mathcal{M}_{n,j-1}^0) \right\|_1 &= \left\| \mathbb{E}(d_{r_n}(Y_{n,2j}, \dot{Y}_{n,2j}) \mid \mathcal{M}_{n,j-1}^0) \right\|_1 \\ &= \tau_1(\mathcal{M}_{n,j-1}^0, Y_{n,2j}) \leq r_n \tau_{1,n}(r_n). \quad \square \end{aligned} \quad (46)$$

**Lemma 3 (Coupling: the sub-blocks)** *Suppose that the random variables  $(X_{n,i})_{1 \leq i \leq n}$  are such (B.2) holds. Moreover, if that for each  $j \in \{2, \dots, \lfloor m_n/2 \rfloor\}$  there is a random variable  $U_{k,j}$  uniformly distributed on  $[0, 1]$  and independent of the  $\sigma$ -algebras  $\mathcal{M}_{n,j-1} = \sigma(Y_{n,1}^{[r_n-l_n]}, \dots, Y_{n,j-1}^{[r_n-l_n]})$*

and  $\sigma(Y_{n,j}^{[r_n-l_n]})$ . Then there exists a random block  $\dot{Y}_{n,j}^{[r_n-l_n]}$ , measurable with respect to  $\mathcal{M}_{n,j-1} \vee \sigma(Y_{n,j}^{[r_n-l_n]}) \vee \sigma(U_j)$ , independent of  $\mathcal{M}_{n,j-1}$  and distributed as  $Y_{n,j}^{[r_n-l_n]}$  such that

$$\left\| \mathbb{E} \left( d_{\cup} (Y_{n,j}^{[r_n-l_n]}, \dot{Y}_{n,j}^{[r_n-l_n]}) \right) \middle| \mathcal{M}_{n,j-1} \right\|_1 \leq r_n \tau_{1,n}(l_n). \quad (47)$$

Moreover, if we set  $\mathcal{M}_{n,j-1} = \sigma(Y_{n,1}^{[r_n-l_n]}, \dots, Y_{n,j-1}^{[r_n-l_n]})$  then the blocks  $(\dot{Y}_{n,j}^{[r_n-l_n]})_{1 \leq j \leq m_n}$  are independent.

*Proof.* The same argument previous proof. However note that the sub-blocks  $(Y_{n,j}^{[r_n-l_n]})_{1 \leq j \leq m_n}$  are separated by  $l_n$  variables.  $\square$

### A.3. Proof of Theorem 1

Let  $(Y_{n,j})_{1 \leq j \leq m_n}$  be the blocks built from  $(X_{n,i})_{1 \leq i \leq n}$ . For  $k \in \{0, 1\}$ , we consider the independent blocks  $(\dot{Y}_{n,2j-k})_{1 \leq j \leq [m_n/2]}$  coupled to the original blocks  $(Y_{n,2j-k})_{1 \leq j \leq [m_n/2]}$ , from Lemma 2. Therefore, if we define  $\Delta_{n,j}^* := f(\dot{Y}_{n,j}) - f(\dot{Y}_{n,j}^{(r_n-l_n)})$ , for  $j = 1, \dots, m_n$ , we have that  $\Delta_{n,j}^*(f) \stackrel{\mathcal{D}}{=} \Delta_{n,j}(f) \stackrel{\mathcal{D}}{=} \Delta_n(f)$ , for each  $j$ , where  $\Delta_{n,j}(f) := f(Y_{n,j}) - f(Y_{n,j}^{(r_n-l_n)})$  and  $\Delta_n(f)$  is defined in (11). Now, if we consider the assumption (C.1), we can apply Petrov [1975]'s Theorem 1 (Section IX.1) to the i.i.d.r.v's  $X_{n,j} := (nv_n)^{-1/2} \Delta_{n,j}^*(f)$ , so

$$D\dot{Z}_n^{(k)}(f) := \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{[m_n/2]} \left( \Delta_{n,2j-k}^*(f) - \mathbb{E} \Delta_{n,2j-k}^*(f) \right) = o_P(1) \quad (48)$$

for  $k = 0, 1$ . In consequence,

$$DZ_n(f) := \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{m_n} (\Delta_{n,j}(f) - \mathbb{E} \Delta_{n,j}(f)) = o_P(1) \quad (49)$$

On the other hand, by Lemma 3, we have that

$$BZ_n(f) := \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{m_n} \left( f(Y_{n,j}^{(r_n-l_n)}) - \mathbb{E} f(Y_{n,j}^{(r_n-l_n)}) \right) \quad (50)$$

converge weakly in fidis if, and only if

$$B\dot{Z}_n(f) := \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{m_n} \left( f(\dot{Y}_{n,j}^{(r_n-l_n)}) - \mathbb{E} f(\dot{Y}_{n,j}^{(r_n-l_n)}) \right) \quad (51)$$

converge weakly in fidis (and in this case the limit distributions are the same). The latter holds because  $B\dot{Z}_n(f) = \dot{Z}_n(f) - D\dot{Z}_n(f)$  and from the assumptions (C.2) and (C.3), where

$$\dot{Z}_n(f) := \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{m_n} (f(\dot{Y}_{n,j}) - \mathbb{E} f(\dot{Y}_{n,j})).$$

Finally, as  $Z_n(f) = BZ_n(f) + DZ_n(f) \forall f \in \mathcal{F}$ , we get the result.  $\square$

#### A.4. Proof of Proposition 1

Suffices to prove the following multidimensional version of Segers [2003]'s condition (6):

$$\lim_{l \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ Y_n^{[l+1:r_n]} \neq 0 | X_{n,1} \neq 0 \right\} = 0, \quad (52)$$

since the rest of the proof follows the same steps of the proof of Drees & Rootzén [2010]'s Lemma 5.2.

Indeed, let  $h(\cdot) = \mathbb{1}\{\cdot \neq 0\}$  be a function defined on  $E^{r_n-l}$ . Consider a increasing sequence of functions  $h_k(\cdot) : E^{r_n-l} \rightarrow [0, 1]$  which approximate to  $h$ , and such that  $\text{Lip}(h_k) = v_k^{-\alpha}$  for some  $\alpha > 0$ . Of course, we set  $k = k(n) \ll n$ . Then,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P} \{ Y_n^{(l+1:r_n)} \neq 0 | X_{n,1} \neq 0 \} &= \limsup_{n \rightarrow \infty} v_n^{-1} \mathbb{P} \{ Y_n^{(l+1:r_n)} \neq 0, X_{n,1} \neq 0 \} \\ &= \limsup_{n \rightarrow \infty} v_n^{-1} \int_{\{X_{n,1} \neq 0\}} \mathbb{P} \{ Y_n^{(l+1:r_n)} \neq 0 | \sigma(X_{n,1}) \} d\mathbb{P} \\ &= \limsup_{n \rightarrow \infty} v_n^{-1} \int_{\{X_{n,1} \neq 0\}} \mathbb{E} [\mathbb{1}\{Y_n^{(l+1:r_n)} \neq 0\} | \sigma(X_{n,1})] d\mathbb{P} \\ &= \limsup_{k \rightarrow \infty} \frac{v_n^{-1}}{v_k^\alpha} \int_{\{X_{n,1} \neq 0\}} \mathbb{E} \left[ \frac{h_k(Y_n^{[l+1:r_n]})}{v_k^{-\alpha}} | \sigma(X_{n,1}) \right] d\mathbb{P} \\ &\leq \limsup_{k \rightarrow \infty} \frac{1}{v_k^{\alpha+1/p}} \left\| \mathbb{E} \left[ \frac{h_k(Y_n^{[l+1:r_n]})}{v_k^{-\alpha}} | \sigma(X_{n,1}) \right] \right\|_p \\ &\leq \limsup_{k \rightarrow \infty} \left[ \frac{(r_n - l)}{v_k^{\alpha+1/p}} \tau_{n,p}(l) + (r_n - l) v_n^{1/q} \right] \leq \limsup_{k \rightarrow \infty} \frac{r_n \tau_{n,p}(l)}{v_k^{\alpha+1/p}}. \end{aligned}$$

Finally, taking  $l \rightarrow \infty$  we have the limit (52) proven.  $\square$

#### A.5. Proof of Theorem 2

Note that  $Z_n$  is asymptotically tight iff  $\dot{Z}_n^{(k)}$  defined by:

$$Z_n^{(k)}(f) := \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{[m_n/2]} (f(Y_{n,2j-k}) - \mathbb{E}f(Y_{n,2j-k})) \quad (53)$$

is asymptotically tight for each  $k \in \{0, 1\}$ . On the other hand, for each  $k \in \{0, 1\}$  we use Lemma 2 together with (B.2) and (D.1) conditions to build independent blocks  $(\dot{Y}_{n,2j-k})_{1 \leq j \leq [m_n/2]}$  coupled to the original blocks  $(Y_{n,2j-k})_{1 \leq j \leq [m_n/2]}$ . In this manner we have that  $Z_n^{(k)}$  is asymptotically tight iff

$$\dot{Z}_n^{(k)}(f) := \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{[m_n/2]} (f(\dot{Y}_{n,2j-k}) - \mathbb{E}f(\dot{Y}_{n,2j-k})) \quad (54)$$

is asymptotically tight, for each  $k \in \{0, 1\}$ . The latter is true due to Theorem 2.11.9 in [Van Der Vaart & Wellner, 1996] by setting  $Z_{nj}(f) = f(Y_{n,j})$  and  $\lfloor m_n/2 \rfloor$  instead of  $m_n$ .

For the remaining assertion we use Theorem 1.  $\square$

#### A.6. Proof of Theorem 3

Consider (T.5). Note that from the triangle inequality  $Z_n$  is asymptotically equicontinuous if  $\dot{Z}_n^{(k)}$  from eqn. (53) is asymptotically equicontinuous for each  $k \in \{0, 1\}$ . Now, again we use Lemma 2 together with (B.2) and (D.1) conditions as in the previous proof to prove that  $Z_n^{(k)}$  is asymptotically equicontinuous iff  $\dot{Z}_n^{(k)}$  is asymptotically equicontinuous for each  $k \in \{0, 1\}$ . However, in this case  $\dot{Z}_n^{(k)}$  is asymptotically equicontinuous from Theorem 2.11.1 in [Van Der Vaart & Wellner, 1996].

The remaining steps are the same of Theorem 2.10's proof in [Drees & Rootzén, 2010].  $\square$

#### A.7. Proof of Proposition 2

The steps are the same that in the proof of Theorem 2.1 in [Drees, 2011], but replacing the assumptions (C1) and (C2) of his paper by our assumptions (B.2) and (B.1), respectively.  $\square$

#### A.8. Proof of Corollary 3

Suffices to replace the assumptions (C1) and (C2) in the proof of Drees [2011]'s Corollary 2.3 by our assumptions (B.2) and (B.1) respectively.  $\square$

#### A.9. Proof of the expression (26)

If  $(X_i)_{i \geq 0}$  is the AR(1)-process (2), note that for each  $i \in \mathbb{N}$

$$X_i = \frac{X_0}{b^i} + \sum_{s=1}^i \frac{\xi_s}{b^{i-s+1}}. \quad (55)$$

Therefore, if  $n$  is sufficiently large such that  $b^d v_n < 1$ , then for  $x \in [0, 1]^d$ :

$$\begin{aligned} \mathbb{P}\{X_{n,1} > x, X_{n,1} \neq 0\} &= \mathbb{P}\{X_i > a_n x_i + u_n, \text{ for some } i = 1, \dots, d\} \\ &= \mathbb{P}\left\{X_0 > b^i(a_n x_i + u_n) - \sum_{s=1}^i b^{s-1} \xi_s, \text{ for some } i = 1, \dots, d\right\} \end{aligned}$$

$$\begin{aligned}
 &= \mathbb{P} \left\{ X_0 > \min_{i=1,\dots,d} \left\{ b^i(a_n x_i + u_n) - \sum_{s=1}^i b^{s-1} \zeta_s \right\} \right\} \\
 &= \sum_{j_1, \dots, j_d \in U(b)} \mathbb{P} \left\{ X_0 > \min_{i=1,\dots,d} \left\{ b^i(a_n x_i + u_n) - \sum_{s=1}^i b^{s-1} j_s \right\}, (\zeta_1, \dots, \zeta_d) = (j_1, \dots, j_d) \right\} \\
 &= \frac{1}{b^d} \sum_{j_1, \dots, j_d \in U(b)} \mathbb{P} \left\{ X_0 > \min_{i=1,\dots,d} \left\{ b^i(a_n x_i + u_n) - \sum_{s=1}^i b^{s-1} j_s \right\} \right\} \\
 &= \frac{1}{b^d} \sum_{j_1, \dots, j_d \in U(b)} \left( \max_{i=1,\dots,d} \left\{ 1 - b^i + \sum_{s=1}^i b^{s-1} j_s + b^i v_n (1 - x_i) \right\} \wedge 1 \right) \\
 &= \frac{1}{b^d} \max_{i=1,\dots,d} \left\{ b^i v_n (1 - x_i) \right\},
 \end{aligned}$$

since  $\mu_b(j_1, \dots, j_d; i) := 1 - b^i + \sum_{s=1}^i b^{s-1} j_s \leq -1$  for all  $(j_1, \dots, j_d) \in U^d(b) \setminus \{(b-1, \dots, b-1)\}$  and  $\mu_b(b-1, b-1, \dots, b-1) = 0$ . Thus,

$$\mathbb{P}\{X_{n,1} > x | X_{n,1} \neq 0\} \xrightarrow{n \rightarrow \infty} \max_{i=1,\dots,d} \{b^{i-d}(1 - x_i)\}. \quad (56)$$

□

#### A.10. Proof of the expression (28)

Let  $x, y \in [0, 1]^d$ . Then as before for  $i \geq 1$ , if  $n$  is sufficiently large such that  $b^{i+d} v_n < 1$ , then we have:

$$\begin{aligned}
 &\mathbb{P}\{X_{n,1} > x, X_{n,i+1} > y\} \\
 &= \mathbb{P} \left\{ X_k > a_n x_k + u_n, X_{i+l} > a_n y_l + u_n, \text{ for some } (k, l) \in \{1, \dots, d\}^2 \right\} \\
 &= \mathbb{P} \left\{ X_0 > \min_{k=1,\dots,d} \left\{ b^k(a_n x_k + u_n) - \sum_{s=1}^k b^{s-1} \zeta_s \right\}, \right. \\
 &\quad \left. X_i > \min_{l=1,\dots,d} \left\{ b^l(a_n y_l + u_n) - \sum_{s=1}^l b^{s-1} \zeta_{s+i} \right\} \right\} \\
 &= \sum_{\substack{j_1, \dots, j_d \in U(b) \\ j_{i+1}, \dots, j_{i+d} \in U(b)}} \mathbb{P} \left\{ X_0 > \min_{k=1,\dots,d} \left\{ b^k(a_n x_k + u_n) - \sum_{s=1}^k b^{s-1} j_s \right\}, \right. \\
 &\quad \left. X_i > \min_{l=1,\dots,d} \left\{ b^l(a_n y_l + u_n) - \sum_{s=1}^l b^{s-1} j_{s+i} \right\}, \right. \\
 &\quad \left. (\zeta_1, \dots, \zeta_d, \zeta_{i+1}, \dots, \zeta_{i+d}) = (j_1, \dots, j_d, j_{i+1}, \dots, j_{i+d}) \right\}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{b^d} \sum_{\substack{j_1, \dots, j_d \in U(b) \\ j_{i+1}, \dots, j_{i+d} \in U(b)}} \mathbb{P} \left\{ X_0 > \min_{k=1, \dots, d} \left\{ b^k(a_n x_k + u_n) - \sum_{s=1}^k b^{s-1} j_s \right\}, \right. \\
 &\quad \left. X_i > \min_{l=1, \dots, d} \left\{ b^l(a_n y_l + u_n) - \sum_{s=1}^l b^{s-1} j_{s+i} \right\}, (\xi_1, \dots, \xi_d) = (j_1, \dots, j_d) \right\} \\
 &= \frac{1}{b^d} \sum_{\substack{j_1, \dots, j_d \in U(b) \\ j_{i+1}, \dots, j_{i+d} \in U(b)}} \mathbb{P} \left\{ X_0 > 1 - \max_{k=1, \dots, d} \left\{ \mu_b(j_1, \dots, j_d; k) + b^k v_n(1 - x_k) \right\}_+ \wedge 1, \right. \\
 &\quad \left. X_i > 1 - \max_{l=1, \dots, d} \left\{ \mu_b(j_{i+1}, \dots, j_{i+d}) + b^l v_n(1 - y_l) \right\}_+ \wedge 1, (\xi_1, \dots, \xi_d) = (j_1, \dots, j_d) \right\} \\
 &= \frac{1}{b^d} \mathbb{P} \left\{ X_0 > 1 - \max_{k=1, \dots, d} \left\{ b^k v_n(1 - x_k) \right\}, \right. \\
 &\quad \left. X_i > 1 - \max_{l=1, \dots, d} \left\{ b^l v_n(1 - y_l) \right\}, \xi_1 = \dots = \xi_d = b - 1 \right\}
 \end{aligned}$$

since  $\mu_b(j_1, \dots, j_d; i) := 1 - b^i + \sum_{s=1}^i b^{s-1} j_s \leq -1$  for all  $(j_1, \dots, j_d) \in U^d(b) \setminus \{(b-1, \dots, b-1)\}$  and  $\mu_b(b-1, b-1, \dots, b-1) = 0$ .

Moreover, note that if  $i > d$

$$\begin{aligned}
 \mathbb{P}\{X_{n,1} > x, X_{n,i+1} > y\} &= \frac{1}{b^d} \mathbb{P} \left\{ X_0 > 1 - \max_{k=1, \dots, d} \left\{ b^k v_n(1 - x_k) \right\}, \right. \\
 &\quad \left. X_i > 1 - \max_{l=1, \dots, d} \left\{ b^l v_n(1 - y_l) \right\}, \xi_1 = \dots = \xi_d = b - 1 \right\} \\
 &= \frac{1}{b^d} \mathbb{P} \left\{ X_0 > 1 - \max_{k=1, \dots, d} \left\{ b^k v_n(1 - x_k) \right\}, \right. \\
 &\quad \left. X_0 > b^i - b^i \max_{l=1, \dots, d} \left\{ b^l v_n(1 - y_l) \right\} + 1 - b^d - \sum_{s=d+1}^i b^{s-1} \xi_s \right\} \\
 &= \frac{1}{b^d} \sum_{j_{d+1}, \dots, j_i \in U(b)} \mathbb{P} \left\{ X_0 > 1 - \max_{k=1, \dots, d} \left\{ b^k v_n(1 - x_k) \right\}, \right. \\
 &\quad \left. X_0 > b^i - b^i \max_{l=1, \dots, d} \left\{ b^l v_n(1 - y_l) \right\} + 1 - b^d - \sum_{s=d+1}^i b^{s-1} j_s, (\xi_{d+1}, \dots, \xi_i) = (j_{d+1}, \dots, j_i) \right\} \\
 &= \frac{1}{b^i} \sum_{j_{d+1}, \dots, j_i \in U(b)} \mathbb{P} \left\{ X_0 > 1 - \max_{k=1, \dots, d} \left\{ b^k v_n(1 - x_k) \right\}, \right. \\
 &\quad \left. X_0 > b^i - b^i \max_{l=1, \dots, d} \left\{ b^l v_n(1 - y_l) \right\} + 1 - b^d - \sum_{s=d+1}^i b^{s-1} j_s \right\}
 \end{aligned}$$



$$\begin{aligned}
 &= \frac{1}{b^i} \sum_{j_{d+1}, \dots, j_i \in U(b)} \min \left( \max_{k=1, \dots, d} \left\{ b^k v_n (1 - x_k) \right\}, \right. \\
 &\quad \left. \max_{l=1, \dots, d} \left\{ b^d + \sum_{s=d+1}^i b^{s-1} j_s + b^{l+i} v_n (1 - y_l) - b^i \right\}_+, 1 \right) \\
 &= \frac{v_n}{b^i} \min \left( \max_{k=1, \dots, d} \left\{ b^k (1 - x_k) \right\}, \max_{l=1, \dots, d} \left\{ b^{l+i} (1 - y_l) \right\} \right) = v_n H_{b,i}(x, y)
 \end{aligned}$$

Similarly for  $1 \leq i < d$ , we obtain that

$$\begin{aligned}
 &\mathbb{P}\{X_{n,1} > x, X_{n,i+1} > y\} \\
 &= \frac{v_n}{b^i} \min \left( \max_{k=1, \dots, i} \{b^k (1 - x_k)\}, \max_{k=i+1, \dots, d} \{b^k \min(1 - x_k, 1 - y_k)\}, \max_{k=d-i, \dots, d} \{b^{k+i} (1 - y_k)\} \right) \\
 &= v_n H_{b,i}(x, y)
 \end{aligned}$$

From Lemma 5.2 - (iii) in [Drees & Rootzén, 2010],  $\mathbb{E}|f(Y_n)| = o(\sqrt{nv_n})$ . Thus, for  $n$  sufficiently large:

$$\begin{aligned}
 &\frac{\text{Cov}(f_x(Y_n), f_y(Y_n))}{r_n v_n} \sim \mathbb{P}\{X_{n,1} > x, X_{n,1} > y\} \\
 &\quad + \sum_{i=1}^{r_n-1} \left(1 - \frac{i}{r_n}\right) (\mathbb{P}\{X_{n,1} > x, X_{n,i+1} > y\} + \mathbb{P}\{X_{n,1} > y, X_{n,i+1} > x\}) \\
 &\xrightarrow{n \rightarrow \infty} \min \left( \max_{k=1, \dots, d} \{b^k (1 - x_k)\}, \max_{k=1, \dots, d} \{b^k (1 - y_k)\} \right) + \sum_{i=1}^{\infty} (H_{b,i}(x, y) + H_{b,i}(y, x)).
 \end{aligned}$$

□

#### A.11. Proof of the expression (31)

The proof is similar to the proof of the expression (26). Indeed,

$$\begin{aligned}
 &\mathbb{P} \left\{ \max_{1 \leq i \leq r_n} X_i > 1 - v_n t \right\} = \mathbb{P} \{X_i > 1 - v_n t, \text{ for some } i = 1, \dots, r_n\} \\
 &= \mathbb{P} \left\{ X_0 > b^i (1 - v_n t) - \sum_{s=1}^i b^{s-1} \xi_s, \text{ for some } i = 1, \dots, r_n \right\} \\
 &= \mathbb{P} \left\{ X_0 > \min_{1 \leq i \leq r_n} \left\{ b^i (1 - v_n t) - \sum_{s=1}^i b^{s-1} \xi_s \right\} \right\}
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j_1, \dots, j_{r_n} \in U(b)} \mathbb{P} \left\{ X_0 > \min_{1 \leq i \leq r_n} \left\{ b^i(1 - v_n t) - \sum_{s=1}^i b^{s-1} \xi_s \right\}, (\xi_1, \dots, \xi_{r_n}) = (j_1, \dots, j_{r_n}) \right\} \\
 &= \frac{1}{b^{r_n}} \sum_{j_1, \dots, j_{r_n} \in U(b)} \mathbb{P} \left\{ X_0 > \min_{1 \leq i \leq r_n} \left\{ b^i(1 - v_n t) - \sum_{s=1}^i b^{s-1} j_s \right\} \right\} \\
 &= \frac{1}{b^{r_n}} \sum_{j_1, \dots, j_{r_n} \in U(b)} \min \left( \max_{1 \leq i \leq r_n} \left\{ 1 + \sum_{s=1}^i b^{s-1} j_s - b^i(1 - v_n t) \right\}, 1 \right).
 \end{aligned}$$

□

**Acknowledgements.** Warm thanks are due to the very constructive and friendly help of Olivier Wintenberger in the redaction and the finalization of this paper.

Very specials thanks are also due to an anonymous referee who pointed clearly the weaknesses of a previous version of this work and helped us to make it more adequate for publication.

## REFERENCES

- Andrews, D. *Non strong mixing autoregressive processes*. J. Appl. Probab. 21, 930-934; (1984).
- Ango Nze P. & Doukhan P. *Weak dependence and applications to econometrics*. Econom Theory 20: 995 - 1045; (2004).
- Basrak, B. & Segers, J. *Regularly varying multivariate time series*. Stoch. Proc. Appl. 119, 1055 - 1080; (2009).
- Dedecker, J. & Prieur, C. *Coupling for  $\tau$ -dependent sequences and applications*. Journal of Theor. Probab. 17-4, 861-885; (2004a).
- Dedecker, J., Prieur, C. *Couplage pour la distance minimale*. C. R. Acad. Sci. Paris Serie 1 338-10, 805-808; (2004)
- Dedecker, J., Prieur, C. *New dependence coefficients. Examples and applications to statistics*. Prob. Theor. and Rel. Fields 132, 203-236; (2005).
- Dedecker, J., Doukhan, P., Lang, G., León, J.R., Louhichi, S. & Prieur, C. *Weak dependence: With Examples and Applications* Lecture Notes in Statistics 190, Springer-Verlag; (2007).
- Doukhan, P. *Mixing: Properties and Examples..* Lect. Notes Statis. 85; (1994).
- Doukhan, P. & Louhichi, S. *A new weak dependence condition and applications to moment inequalities*. Stoch. Proc. Appl. 84. 313-342; (1999).

- Doukhan, P., Teyssiere, G. and Winant, P. *Vector valued ARCH( $\infty$ ) - processes, in Dependence in Probability and Statistics*, P. Bertail, P. Doukhan and P. Soulier Eds. Lecture Notes in Statistics, Springer, New York (2006).
- Doukhan, P. *Stochastic Models for Time Series*. Springer-Verlag; (2015).
- Drees, H. *Weighted Approximations of Tail Processes for  $\beta$ -Mixing Random Variables*. Ann. Appl. Probab. 10, 1274-1301; (2000).
- Drees, H. *Tail empirical processes under mixing conditions*. In: H.G. Dehling, T. Mikosch and M. Sorensen (eds.), *Empirical Process Techniques for Dependent Data*, 325-342. Birkhuser, Boston; (2002).
- Drees, H. *Extreme Quantile Estimation for Dependent Data with Applications to Finance*. Bernoulli 9, 617-657; (2003).
- Drees, H & Rootzén, H. *Limit Theorems for Empirical Processes of Cluster Functionals*. Ann. Stat. 4, 2145 - 2186; (2010).
- Drees, H. *Bias correction for blocks estimators of the extremal index*. Univ. Hamburg; (2011).
- Eberlein, E. *Weak convergence of partial sums of absolutely regular sequences*. Statist. Probab. Letters 2, 291 - 293; (1984).
- Gómez, J.G. *Dependent Lindeberg CLT - Finite Dimensional for Empirical Processes of Cluster Functionals*. Preprint in arXiv:1404.4989; (2015).
- Perfekt, R. *Extremal behaviour of stationary Markov chains with applications*. Ann. Appl. Prob. 4, 529 - 548; (1994).
- Petrov, V.V. *Sums of Independent Random Variables*. Springer, Berlin; (1975).
- Pickands, J. *Statistical inference using extreme order statistics*. Ann. Statist. 3, 119-131; (1975).
- The tail empirical process for stationary sequences*. Technical report, Department of Mathematics, Chalmers University, Sweden; (1995).
- Rootzén, H. *Weak convergence of the tail empirical function for dependent sequences*. Stoch. Proc. Appl. 119, 468 - 490; (2009).
- Segers, J. *Functionals of clusters of extremes*. Adv. Appl. Probab. 35, 1028 - 1045; (2003).
- Yun, S. *The extremal index of a higher-order stationary Markov chain*. Ann. Appl. Prob. 8, 408 - 437; (1998).
- Yun, S. *The distribution of cluster functionals of extreme events in a  $d$ th-order Markov chain*. J. Appl. Probab. 37, 29 - 44; (2000).
- Van Der Vaart, A.W. and Wellner, J. A. *Weak Convergence and Empirical Processes*. Springer, New York; (1996).